ANTHROP\C

June 10, 2026

The Honorable Tim Scott
Chairman
Committee on Banking, Housing, and Urban Affairs
534 Dirksen Senate Office Building
Washington, DC 20510

The Honorable Elizabeth Warren
Ranking Member
Committee on Banking, Housing, and Urban Affairs
534 Dirksen Senate Office Building
Washington, DC 20510

RE: Illicit access to American AI models by Alibaba-affiliated operators

Dear Chairman Scott and Ranking Member Warren:

Alibaba executed the largest known distillation attack on Anthropic to date. We write ahead of the Committee's June 11 hearing, "AI and the American Dream," with new, confidential evidence of the largest campaign to illicitly extract Claude's capabilities we have ever measured, conducted by operators affiliated with Alibaba and Alibaba Qwen, Alibaba's AI lab. Between April 22 and June 5, 2026, this campaign generated more than 28.8 million exchanges with Claude through almost 25,000 fraudulent accounts, in violation of our terms of service and access restrictions. Alibaba's campaign targeted some of Claude's most valuable capabilities, such as agentic reasoning, software engineering, and long-horizon tasks. Congress should advance measures that facilitate threat information sharing between US AI labs, close loopholes allowing PRC AI labs to access advanced US chips, and penalize PRC labs responsible for distillation attacks.

Distillation attacks harvest American IP to advance our competitors. This campaign followed many of the same patterns we disclosed in a public blog post in February spotlighting the distillation attacks of DeepSeek, Moonshot, and Minimax, which together generated over 16 million exchanges with Claude through 24,000 fraudulent accounts. These distillation attacks are carried out illicitly, systematically, and at industrial scale to harvest US AI capabilities across frontier labs and repackage them as their own without incurring the training and R&D costs required to train US frontier models.

Distillation attacks help China reach Mythos Preview-level capabilities sooner. The release of Mythos Preview underscored the importance of ensuring a durable lead in AI to safeguard US national security. Had China achieved a Mythos-caliber model before the United States, the PRC would have possessed advanced cyber capabilities to deploy against the US government and American companies and exploit vulnerabilities faster than previously possible. The larger the capability gap is between US and PRC AI models, the more time the US government will have to harden cyber defenses and adopt AI systems across national security domains, as we have seen with Project Glasswing, which has made key agencies and companies less vulnerable to cyber intrusions. Furthermore, PRC AI labs tend to invest fewer resources than their US counterparts in ensuring models are deployed responsibly. PRC AI models are

often released with weak safeguards that are easily jailbroken, enabling other US adversaries to use these models for a wide range of activities that run contrary to US interests.

Alibaba brazenly ran this campaign *after* the Trump Administration's OSTP memo on curbing distillation attacks. Beyond its scale, this campaign was striking for its brazen nature. Alibaba is listed on the New York Stock Exchange, maintains business operations in the United States, and is accountable to US investors and regulators. Yet this activity unfolded in the weeks after the OSTP NSTM-4 memorandum, in which the Trump Administration found that PRC-based entities were conducting industrial-scale campaigns to distill US frontier AI models, calling the behavior "unacceptable." In proceeding with these distillation attacks, Alibaba ignored the Trump Administration's warnings.

Alibaba's behavior advances PLA AI efforts. The stakes here extend well beyond a single company's intellectual property. Just this week, the Pentagon added Alibaba to the 1260H List of "Chinese military companies." As AI becomes a more important enabler of military, intelligence, and cyber capabilities, every successful distillation campaign that shrinks the US-PRC gap through illicit model access erodes the United States' technological edge in those domains. Recent research has shown that AI adoption is a top priority for China's military modernization objectives, and the PLA is already using LLMs developed by leading PRC AI labs across military applications and domains. PRC state actors are weaponizing AI to enable cyber operations. Anthropic, Google, and OpenAI have all published reports on PRC state actors using advanced AI systems to augment and automate cyber operations against US and allied targets.

Distillation attacks turn hundreds of billions of dollars in American investment and R&D into a massive subsidy for our geopolitical competitors. When PRC labs distill these capabilities from US models, they capture the returns on American investments without bearing the costs or risks associated with training frontier AI models. This inverts the economic logic that underwrites American AI leadership, turning billions of dollars worth of research and development, compute, and other US investments into a subsidy for our competitors. US firms shoulder the full expense of advancing the frontier, while PRC competitors free-ride on the result and bring near-equivalent capabilities to market.

Despite a strong response from the US government, distillation attacks remain widespread practice among PRC AI labs. Anthropic is supportive of the US government's efforts to combat these distillation attacks. These include NSPM-11, which directs the national security enterprise to partner with private-sector AI companies to help secure American AI from distillation attacks, including through threat-intelligence sharing and joint red-team exercises, and NSTM-4, which found that foreign entities based in the PRC are conducting industrial-scale campaigns to distill US frontier AI systems and committed the executive branch to information sharing with industry, joint development of defensive best practices, and consideration of measures to hold responsible foreign actors accountable.

Still, these whole-of-government efforts to counter distillation attacks have not dissuaded PRC AI labs' attempts to gain access to US frontier AI capabilities. While Alibaba's distillation campaign is notable for its scope and scale, orchestrating sophisticated, industrial-scale distillation attacks has become a

widespread practice for PRC AI labs. To evade detection, PRC AI labs access US AI models using obfuscation techniques and proxy networks, supplied by a growing circumvention economy.

Protect US labs's ability to share intel and counter these adversarial operations. More action is needed to ensure continued American AI leadership. First, Congress should advance legislation that would enable deeper collaboration to combat distillation attacks, both between the US government and leading frontier labs as well as between the frontier labs themselves. Congress should clarify antitrust guidelines to allow AI labs to share information about tactics used by PRC AI labs that engage in distillation attacks, enabling more effective detection and prevention of these activities.

Export controls on advanced American compute are critically important to curbing PRC distillation attacks. Second, Congress should codify efforts to limit the PRC's ability to access advanced US compute, a critical input to the training of AI models. Distillation relies on compute—without ample compute, PRC AI labs would not be able to as effectively use the harvested exchanges to achieve advanced capabilities. Today, PRC AI labs continue to be able to access American AI chips through illicit and evasive means, including through smuggling networks and by accessing data centers outside of China's borders. We encourage Congress to advance legislation that would close these loopholes, thereby hamstringing PRC AI labs' ability to use distillation attacks effectively.

Penalize bad behavior from the PRC labs. Third, Congress should pass legislation that would make it more difficult and costly for PRC AI labs to access US AI models and launch distillation attacks. Anthropic does not allow commercial access to Claude for entities in China and for subsidiaries of PRC-headquartered companies. This policy reflects the reality that firms under PRC jurisdiction are often required or otherwise compelled to advance PRC state objectives, including those that run counter to US national security interests. Accordingly, Anthropic supports efforts that would leverage the US economic security toolkit to penalize PRC AI labs responsible for conducting distillation campaigns in order to deter this behavior, including measures that would limit those actors' ability to use US AI models, acquire advanced US chips, and access data centers located outside China.

Sincerely,

Sarah Heck
Head of Policy, Anthropic